Chapter 3
The Need to Build Trust

An organization operating and managing a complex adaptive information and communications technology (ICT) system is said to be anti-fragile when, over time, the organization is able to protect the user population from serious consequences of system failures and simultaneously provide digital services fulfilling the users' changing needs [41]. According to Chap. 2, failures are inevitable in a complex ICT system. Unless a user population has a high level of trust in the system, the population may abandon the system after a failure. Hence, any anti-fragile organization running a complex ICT system must maintain a high level of trust over time to keep their users after inevitable system failures.

To better understand why it is critical for any anti-fragile organization to maintain user trust, this chapter first defines the concept of trust and then develops a model of a user population whose individuals influence each others' levels of trust in an ICT system, for example, an e-government platform with digital services. The model demonstrates that a population's trust decreases rapidly when distrust within small groups of individuals starts to spread. Further, it illustrates why it is hard to determine which incidents will lead to widespread distrust and clarifies why it is very difficult to create pervasive trust when there is much distrust. We find that a population's trust is fragile to incidents directly affecting a few individuals while widespread distrust is robust against concentrated efforts to rebuild trust. Finally, the chapter discusses approaches to limit the spread of distrust and maintain a high level of trust.

3.1 Defining Trust

Trust can be viewed as a computational construct whose value depends on the context. The value is likely to change over time. Here, an individual's trust in an entity is specified by three concepts: *trust*, *mistrust*, and *distrust*, viewed as mutually exclusive states representing different degrees of trust. Mistrust represents a general sense of unease toward an ICT system based on mostly unverified information, while users

© The Author(s) 2016

K.J. Hole, *Anti-fragile ICT Systems*, Simula SpringerBriefs on Computing 1,
DOI 10.1007/978-3-319-30070-2_3

distrust a system because of negative experiences or reliable information from experts about serious problems with the system. As an example, users distrust an Internet banking system after suffering financial losses, but they only mistrust the system after being told about security problems by family, friends, or co-workers.

Since most users do not fully understand how an ICT system operates or why incidents occur, they will seek advice from others about what to believe about the system; that is, their levels of trust are influenced by other stakeholders. Mistrust is a less stable state than distrust. While users with mistrust are likely to develop distrust when they receive additional negative information about a system or when they become victims of actual incidents, users harboring distrust are less likely to move back to a state of mistrust because they have already suffered harm caused by the system.

An individual who trusts an entity has a positive expectation of the entity's future behavior [42, 43]. The individual will cooperate with the entity to reach a certain goal, even though it is possible that the entity will misbehave and inflict costs or damage on the individual. The entity gains the individual's trust over time through repeated actions benefiting the individual.

An individual harboring mistrust believes the uncertainty is too large to expect a particular behavior from an entity. A citizen may, for example, believe in the government's sincere desire to deliver highly secure services on the Web, but has little or no confidence in the government's ability to actually deliver adequate security.

An individual distrusting an entity believes the entity will deliberately act against him or her in a given situation. A citizen harboring distrust may think that the government intentionally overstates the security of its e-government services or uses collected personal information to spy on individuals.

While a citizen's trust in a system can be in one of only three states in this chapter, the whole population has different degrees of trust, mistrust, and distrust at the same time, measured by the fractions of individuals in each of the three states. Note that the three fractions sum to one.

To illustrate a population's mistrust and distrust of an ICT system, as well as its owner, we consider a large identity management system that was never fully implemented. A former UK government under Labour started to deploy a centralized identity system, called the National Identity Scheme (NIS), to provide biometric identity cards to all lawful residents aged 16 and over. Roughly £250 million were spent developing NIS (http://news.bbc.co.uk/2/hi/8707355.stm).

The London School of Economics and Political Science started the Identity Project to analyze NIS. Project members mistrusted the UK government, accusing it of not understanding the political, social, and technological risks of establishing a national ID system with a centralized database containing up to 50 data points per individual [44].

Over the years, the Identity Project published reports and participated in the national debate to convince politicians to scrap NIS. The lobby group NO2ID also opposed the creation of NIS. Their briefing papers imply distrust of the UK government. In particular, NO2ID discussed how NIS could allow the government to

manage society by spying on people, severely compromising their privacy and security.

The UK Labour government allowed mistrust and distrust to grow by relegating, ignoring, or attacking independent experts pointing out weaknesses in NIS [44, pp. 81–2], [45]. The predominantly negative press coverage of NIS helped spread mistrust and distrust when people started to discuss it. According to a study of UK newspapers [45], NIS was portrayed as unsafe, lacking accountability, compulsory rather than based on choice, universal, tough on immigration, and creating an imbalance between liberty and security. In 2010, the new Conservative coalition government's Identity Documents Act abolished the identity cards and ordered the destruction of all data in the associated National Identity Register.

3.2 Explanatory Trust Model

The following discrete-time model provides an explanation for how trust, mistrust, and distrust change in a population due to incidents in a complex ICT system. Patches on a square represent the modeled individuals. The square wraps around at the edges, that is, the model has a doughnut shape. An individual's state of trust is represented by the color of the patch: Trust is green (■), mistrust is yellow (■), and distrust is red (■), as seen in Fig. 3.1. Each individual has eight neighbors. At each time step, the state of an individual is updated based on the states of its neighbors.

Since it is not obvious how to update the patches, we study 14 sets of update rules defined by the columns of Table 3.1. Each set has two rules defining changes from trust to mistrust and from mistrust to distrust, as well as two rules defining changes in the opposite directions. The two first (last) rules induce a color change when the number of green neighbors is no larger (no smaller) than a threshold. To clarify, the four rules defined by the rightmost column in Table 3.1 are as follows:

(i) A green patch changes to yellow when a maximum of four neighbors are green.
(ii) A yellow patch turns red when a maximum of three neighbors are green.
(iii) A red patch turns yellow when a minimum of seven neighbors are green.
(iv) A yellow patch turns green when a minimum of six neighbors are green.

The two first rules in a set repeatedly reduce a population's trust as individuals become increasingly surrounded by individuals with mistrust or distrust. The rules create an escalating feedback loop producing increasingly more mistrusting and distrusting individuals when the initial conditions are right. The two last rules create a dampening feedback loop when the starting conditions are right, but this time to increase the population's trust.

All 14 rule sets defined by the columns in Table 3.1 result in the same change pattern: An individual with trust goes through a period of mistrust before developing distrust and an individual with distrust develops mistrust before trust. Individuals who have trusted an entity for a long time are reluctant to mistrust or distrust it.

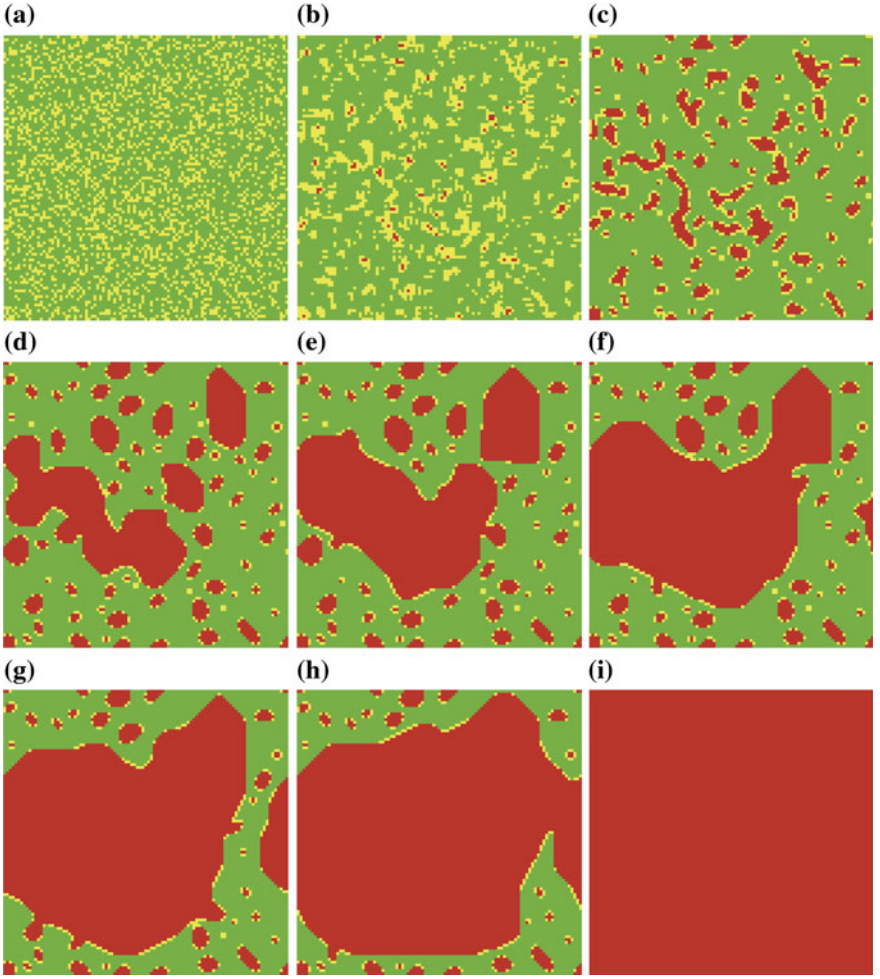

Fig. 3.1 Development of mistrust in a 100×100 population of patches. The initial mistrust is 27 % (*yellow patches*) at time step $t = 0$. The following snapshots show the formation and spreading of distrust (*red patches*) over time. Rules (i)–(iv) were used. **a** $t = 0$, **b** $t = 1$, **c** $t = 3$, **d** $t = 20$, **e** $t = 40$, **f** $t = 60$, **g** $t = 80$, **h** $t = 100$, **i** $t = 165$

Distrusting individuals are even more reluctant to ever again trust an entity that has violated their trust and caused pain or damage. Finally, an individual harboring mistrust develops distrust when surrounded by much mistrust.

Table 3.1 Each column defines a set of four update rules

Changes	Color-changing thresholds													
	Maximum number of green neighbors													
🟩 → 🟨	3	3	3	3	3	3	3	3	4	4	4	4	4	4
🟨 → 🟥	1	1	1	1	2	2	2	2	1	1	2	2	3	3
	Minimum number of green neighbors													
🟥 → 🟨	6	6	7	7	6	6	7	7	6	7	6	7	6	7
🟨 → 🟩	5	6	5	6	5	6	5	6	6	6	6	6	6	6

The two first entries in a column define the maximum number of green neighbors causing changes toward distrust, while the two last entries define the minimum number of green neighbors needed to change away from distrust

3.3 Model Limitations

Since a model is a simplification of a real-world system, it is possible to create many models emphasizing different aspects of the real system. We have introduced a simple model of a population's trust in a system. It is possible to add more functionality to this model. As an example, we could equip the individuals with a memory of past incidents. Furthermore, while all individuals react the same way in the current model, it is possible to use different rules for different individuals. Finally, many other update rules are possible.

Alternatively, we could define a trust model by a graph where the nodes represent individuals and the edges connect nodes that influence each other. When the views of experts and commentators are widely reported by the media, a few nodes have a very large number of edges to neighboring nodes. While our model does not include these "super-spreaders"directly, their combined influence is represented by the initial pattern of mistrust. The more negative the media coverage, the higher the percentage of initial mistrust.

The trust model is non-predictive, in the sense that it cannot forecast a population's trust in a real system. However, it offers an explanation of how the degree of trust changes in a large community of users.

3.4 Trust Is Fragile

We first study how a high degree of trust can turn into a high degree of distrust. We concentrate on system incidents reported in the media. While most incidents go unnoticed by the media, a few incidents are widely reported. Not all reported events are very serious from a technical point of view, but extensive media coverage can still create mistrust among a significant fraction of users.

The explanatory model was implemented in NetLogo [46] and the highlighted rules (i)–(iv) were used to generate the figures. At the start of a model run, a selectable percentage of all individuals is yellow (mistrust) and the remaining percentage is green (trust). The yellow patches are selected at random. Initially there is no distrust. Figure 3.1 shows snapshots of a model run with an initial mistrust of 27 % in a population of 10,000 patches. Figure 3.1a depicts many small localized outbreaks of mistrust at time step $t = 0$ due to widespread media coverage of an incident. Distrust starts to occur already at time step $t = 1$. The distrust forms isolated islands that start to combine as they become larger. The run ends when the patches' color patterns no longer change. At the end of the run in Fig. 3.1i, there is 100 % distrust.

Figure 3.2 plots the final fraction of distrust as a function of the initial fraction of mistrust. Each column in the plot was averaged over 100 runs with the same initial fraction of mistrust. As long as the initial density of mistrust is less than 15 %, the resulting fraction of distrust is less than 1 %, on average. However, around 15 % of initial mistrust, there is a transition where increasing mistrust rapidly results in very large fraction of distrust. An initial mistrust of 28 % results in 99 % distrust, on average. Experiments with the additional 13 leftmost rule sets in Table 3.1 all revealed similar sharp transitions to massive distrust starting at fairly low percentages (16–33 %) of initial mistrust. Since it is difficult to determine when these transitions occur in real systems, it is hard to predict if an incident will lead to massive distrust.

The model indicates (but does not prove) that user trust in a complex ICT system is fragile, because an incident affecting a few users can create massive distrust when extensive media reporting creates enough initial mistrust. The UK Labour government did not handle the media skillfully. Therefore, extensive negative press helped create enough distrust to stop NIS. Of course, an incident affecting many users directly can create enough initial mistrust without any help from the media. According to the explanatory model, both cases result in pervasive mistrust.

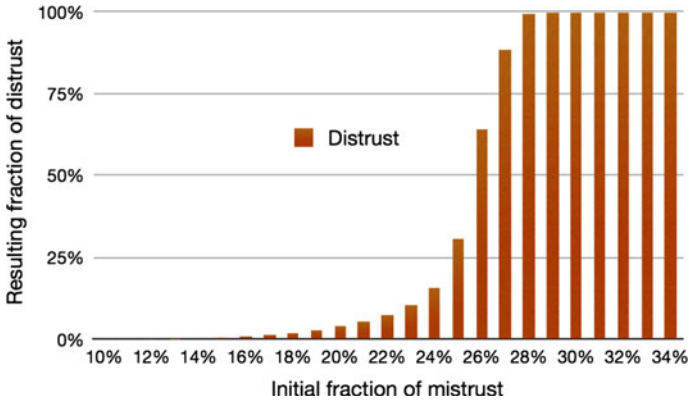

Fig. 3.2 Average fraction of distrust as a function of the initial fraction of mistrust in a population of 100 × 100 patches. A transition starts around 15 %

3.5 Distrust Is Robust

Next, we determine when the model moves from a high percentage of distrust to a high percentage of trust. At the start of a model run, all the patches are red, that is, there is 100 % distrust. A selectable percentage of the red patches chosen at random then change to green as the model starts to run. Figure 3.3 plots the resulting percentage of trust as a function of the initial percentage of green patches, again using the rules (i)–(iv) detailed earlier. Each column of the plot is averaged over 100 runs. There is a rapid transition around 80 % initial trust. Below this transition, the model returns to 100 % distrust. The plot demonstrates how hard it is to create widespread trust when there is massive initial distrust.

Experiments with the 13 additional sets of rules in Table 3.1 also showed similar sharp transitions at large values (42–80 %) of initial trust. The model again returns to 100 % distrust below these transitions. The model implies that massive distrust in a complex ICT system is robust to large efforts to create widespread trust. It will take a sustained effort over a long period to rebuild trust. There is no guarantee that such an effort will succeed. In fact, it may be close to impossible to rebuild widespread trust in a system when there is massive distrust among the user population.

A few comments are needed to fully understand both the limitations and implications of all the reported experiments. While it is unlikely that a large population has 100 % trust or distrust in a real system, it is not unlikely that the population's trust varies sharply, as depicted in Figs. 3.2 and 3.3. However, the experiments do not prove that such transitions exist, especially since we have only explored one of many possible trust models and only deployed a tiny fraction of all possible update rules.

Taken together, the reported experiments suggest that a long-term effort to limit the formation of mistrust should already be started when a system is first created. The effort should be intensified immediately after an incident to avoid a state of massive distrust from which it is very hard to recover. A successful effort to build a good

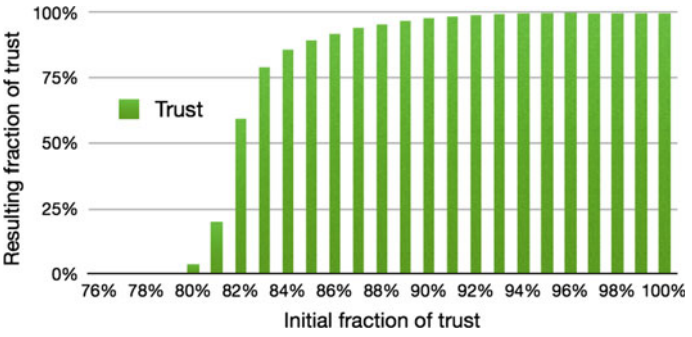

Fig. 3.3 Average fraction of trust as a function of the initial fraction of trust in a population of 100 × 100 patches. A transition starts around 80 %

reputation and to reduce incident reporting translates into a smaller percentage of initial mistrust in the model. As long as the percentage is below the transition point to massive distrust (see Fig. 3.2), the mistrust will die out rather quickly, returning the population to a high level of trust.

3.6 Maintaining Trust

Since it is very hard to recover from massive distrust, an anti-fragile organization has to actively build and maintain its customers' trust. This section first discusses the Tylenol crisis in 1982 to illustrate that it is possible to avoid massive loss of trust, even during very challenging situations. It then discusses specific approaches to build and maintain a user population's trust in an ICT system providing digital services, especially an e-government platform delivering services to an entire nation.

The painkiller Tylenol provided about 15 % of Johnson & Johnson's corporate profit during the first three quarters of 1982. Then somebody laced Tylenol capsules with cyanide and killed seven people in the Chicago area. The company quickly stopped Tylenol production and issued warnings to hospitals and distributors. It then recalled about 30 million Tylenol bottles from the market and advertised in the media to warn people not to use the product. Johnson & Johnson put public safety first, even though the recall was very expensive. The company got much positive press for their resolute handling of the crisis. While Johnson & Johnson's share of the painkiller market fell from around 35–8 % after the killings, the company reintroduced the product and rebounded in less than a year.

3.6.1 Prepare Alternative Services

Whether or not an ICT system is implemented in the cloud, there is always a possibility of a rare, catastrophic incident taking down the system and all its services for a long time. If there are no alternatives to the services offered by an organization, then a long simultaneous failure of all services is intolerable to the organization, because mistrust (followed by distrust) will spread among users, resulting in demands for technical changes and even financial compensation. Consequently, it is a good idea to have alternative solutions to the most important services to reduce the possibility of mistrust and distrust spreading in the user population. A government could for example run its services in a cloud and use another cloud in an emergency. Alternative services should run continuously. If services lie dormant much of the time, there is a significant chance they will not work when needed. For example, it is not uncommon for emergency power systems to not work because they have not been tested for a long time.

According to Geer [31], it is important to retain pre-Internet systems because they have few external dependencies and avoid common mode failures with Internet-based systems. The dismantling of old systems and procedures that have worked well for decades may have serious unintended consequences. National institutions that no longer accept communication on paper exclude a small but significant percentage of the population. Furthermore, states relying solely on electronic voting cannot fall back on traditional paper voting should the electronic voting solutions fail due to technical problems or targeted attacks. Finally, citizens and first responders in countries dismantling their fixed-line phone systems cannot communicate when the mobile phone systems are down. Much of the costs companies and governments save by eliminating redundant systems may be lost when swan incidents take down their remaining unique systems. While it makes sense to eliminate a redundant system in the short run, it can turn out to be a very bad decision in the long run.

3.6.2 Make Digital Services Voluntary

It may be tempting for an organization, especially a government, to "force" individuals to use its digital services. A government can even create a legal obligation to use e-government services to ensure large resource savings. However, the mandatory use of digital services is likely to create mistrust or even distrust because users have little or no control over an organization's actions. Furthermore, some individuals lack the computer skills needed to use the services and others have disabilities forcing them to depend on the help from others. Consequently, it should be possible to opt out of any service without undue difficulty to avoid mistrust and distrust among individuals. In summary, an obligation to use a system leads to mistrust or even distrust, while voluntary use ensures that nearly all new users will trust the system because without trust they will not use it. Since a high fraction of initial trust makes it easier to maintain the necessary trust over time, voluntary use is better than mandatory use.

3.6.3 Build a Good Track Record

It is counterproductive for an organization to ignore or hide the fact that events with a negative impact are inevitable in ICT systems of high complexity. It is a particularly bad policy to rely on spin control after incidents have occurred. An organization should, instead, gain trust by creating a good track record from the start of a new service. The dissemination of practical information to users via the Web and the press is a way to build trust.

An organization must demonstrate competence and quickly fix problems when a large incident occurs. If the organization has a good track record, then users are quite forgiving when they are convinced that an incident was caused by a technical problem [42]. Since the loss of trust can be huge when users suspect malicious

intent, an organization must clarify its intentions, especially how it will use and not use personal information, to prevent the rapid deterioration of trust during an incident.

3.7 Discussion and Summary

To build and maintain an anti-fragile ICT system, it is not enough to use the right system design and the best information technologies; it is also necessary to create an organization that learns from mistakes, values openness, and understands the importance of building and maintaining trust relationships with its customers. If the overall level of trust is high and a system failure is due to an understandable human error or a technical glitch, then customers forgive readily, assuming the organization is open about the cause of the failure and shows competence when rectifying the mistake. An organization that downplays incidents, stonewalls journalists, attacks independent commentators and security experts, and displays arrogance toward its customers risks creating massive distrust in the user population. This chapter illustrates that the organization may be unable to recover from such a position, even if it spends large amounts of resources trying to rebuild trust.

Open Access This chapter is distributed under the terms of the Creative Commons Attribution-Noncommercial 2.5 License (http://creativecommons.org/licenses/by-nc/2.5/) which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

The images or other third party material in this chapter are included in the work's Creative Commons license, unless indicated otherwise in the credit line; if such material is not included in the work's Creative Commons license and the respective action is not permitted by statutory regulation, users will need to obtain permission from the license holder to duplicate, adapt or reproduce the material.